

UNITED STATES AIR FORCE RESEARCH LABORATORY

MISSION COMPLEXITY SCORING IN DISTRIBUTED MISSION TRAINING

Todd Denning

L3 Communications, Link Simulation & Training Div
6030 South Kent Street
Mesa AZ 85212-6061

**Winston Bennett, Jr.
Peter Crane**

Air Force Research Laboratory
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

MAY 2003

**AIR FORCE MATERIEL COMMAND
AIR FORCE RESEARCH LABORATORY
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061**

Approved for public release; distribution is unlimited.

NOTICES

Publication of this report does not constitute approval or disapproval of the ideas or findings. It is published in the interest of STINFO exchange.

Using Government drawings, specifications, or other data included in this document for any purpose other than Government-related procurement does not in any way obligate the US Government. The fact that the Government formulated or supplied the drawings, specifications, or other data, does not license the holder or any other person or corporation, or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

WINSTON BENNETT, JR.
Project Scientist

HERBERT H. BELL
Technical Advisor

CURTIS J. PAPKE, Colonel, USAF
Chief, Warfighter Training Research Division

Copies of this report may be requested from:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, VA 22060-6218
<http://stinet.dtic.mil>

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) May 2003		2. REPORT TYPE Final		3. DATES COVERED (From - To) August 2001 to August 2002	
4. TITLE AND SUBTITLE Mission Complexity Scoring in Distributed Mission Training				5a. CONTRACT NUMBER F33615-97-D-5000	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 62205F	
6. AUTHOR(S) Todd Denning Winston Bennett Jr.* Peter Crane*				5d. PROJECT NUMBER 1123	
				5e. TASK NUMBER B0	
				5f. WORK UNIT NUMBER 01	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) L3 Communications Link Simulation & Training Division 6030 South Kent Street Mesa AZ 85212-6061				8. PERFORMING ORGANIZATION REPORT NUMBER *Air Force Research Laboratory Warfighter Training Research Division 6030 South Kent Street Mesa AZ 85212-6061	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory Warfighter Training Research Division 6030 South Kent Street Mesa AZ 85212-6061				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-HE-AZ-TR-2002-0192	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES Air Force Research Laboratory Technical Monitor: Dr Winston Bennett Jr. , AFRL/HEA, 480.988.6561 x-297, DSN 474-6297. Portions of this report were published in the Proceedings of the 2001 Interservice/Industry Training, Simulation, and Education Conference.					
14. ABSTRACT Training specialists and subject-matter experts at the Air Force Research Laboratory's Mesa Research Site have been characterizing Distributed Mission Training (DMT) scenarios in terms of specific learning objectives linked to mission-essential competencies and to the underlying knowledge, skills, and experiences that are required for successful combat performance. As part of scenario characterization, we have identified the mission characteristics and levels of those characteristics that are important for complexity indexing. Techniques for determining overall scenario complexity and for relating scenario characteristics to mission essential competencies have been developed. This report presents results from a validation study comparing the new, analytically based complexity methodology with an empirically based approach. In addition, applications of the new assessment to both F-15 and F-16 weapon system capability and scenario characteristics to learning objectives while controlling overall complexity are discussed together with plans for developing DMT instructor support systems.					
15. SUBJECT TERMS Combat Performance; Distributed Mission Training; DMT; Instructor support systems; Learning; Learning objectives; Mission complexity scoring; Mission-essential competencies;					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UNLIMITED	18. NUMBER OF PAGES 26	19a. NAME OF RESPONSIBLE PERSON Ms Liz Casey
a. REPORT UNCLASSIFIED	b. ABSTRACT UNCLASSIFIED	c. THIS PAGE UNCLASSIFIED			19b. TELEPHONE NUMBER (include area code) 480.988.6561 x- 188 DSN 474-6188

EXECUTIVE SUMMARY

At the December 2001 Interservice/Industry Training, Simulation, and Education Conference (IITSEC) held in Orlando FL (Crane, Robbins, Bennett & Bell, 2001), we presented a first look at an empirically based methodology for quantifying the complexity of Distributed Mission Training (DMT) scenarios for simulation-based, team training. The outcome of that effort was the development of a scenario complexity index derived from subject-matter expert judgments which has been used to assess the effects of training in a building-block syllabus. Over the past few months, training specialists and subject-matter experts at Air Force Research Laboratory's Mesa Research Site have been characterizing DMT scenarios in terms of specific learning objectives linked to mission-essential competencies and to the underlying knowledge, skills, and experiences that are required for successful combat performance. As part of scenario characterization, we have identified the mission characteristics and levels of those characteristics that are important for complexity indexing. Techniques for determining overall scenario complexity and for relating scenario characteristics to mission essential competencies have been developed.

This report presents results from a validation study comparing the new, analytically-based complexity methodology with the empirically-based approach presented last year. In addition, applications of the new assessment to both F-15 and F-16 weapon system capability and scenario characteristics to learning objectives while controlling overall complexity will be discussed together with plans for developing DMT instructor support systems.

MISSION COMPLEXITY SCORING IN DISTRIBUTED MISSION TRAINING

INTRODUCTION

Fighter training program development, as currently conducted, is more art than science. In general, fighter training syllabi produced in the field present theoretical credibility; however, scrutiny of practical application frequently reveals detriments in learning pace as a result of poorly staged training stressors within mission events. Existing training theories and papers highlighted in previous complexity research provide varied approaches to achieving competence in combat operations either using, *or specifically avoiding*, learning objectives (Crane, Robbins, Bennett, & Bell, 2001). The debate for and against objective-based training is interesting, but even in highly structured efforts such as Instructional System Development (ISD), theory falls victim to an expedient and less effective practical application in many military syllabi. In a similar outcome, Simulator Networking (SIMNET) studies showed promise in proficiency gains through voluminous unstructured experiences (Alluisi, 1991), yet Crane et al.'s (2001) analysis of SIMNET still recognizes a need for objective-based, building-block training suggested by ISD for primary and resource-limited training evolutions. Whether structured or unstructured, current training formats fall short of the mark in efficiently attaining true mission competence.

A survey of the practicalities of fighter training points toward a prevailing detractor. The root of the problem in program development *is the typical gulf of separation between the knowledge fields of expert combat operators and instructional system designers*. Researchers at Air Force Research Laboratory (AFRL) have introduced methodology to identify competencies, knowledge, skills, and experiences required to achieve combat success (Bennett, Schreiber & Andrews, in press; Colegrove & Alliger, 2002). Through this effort, both researchers and combat operators charged with training development duties will gain better insight into learning processes and how they link to competencies required in the field.

Budgetary stress demands improved training efficiency. Selection of an optimized training path for warfighters has economic and security advantages that, in the current climate, demand more than a dusting off and review of an existing syllabus from another mission or aircraft. Despite USAF's embrace of ISD, the core training program for multirole fighters has survived relatively intact from the F-100 era to the present through a "review and renew" process carried out by teams of combatants-turned-trainers. Combatants are rarely schooled in ISD nor are they able to catalog the complexities resident in many of their chosen training scenarios. In the current budgetary climate, there is no evidence to suggest fiscal limitations will tolerate SIMNET's unstructured approach to proficiency development, particularly for formal training of combatant candidates. A need will continue to exist for structured exposure to warfighting concepts that begins with low complexity and paces the program accurately to higher complexities for maximum training gains in short-term formal courses such as the F-16 Basic Qualification Course and the F-16 Weapons Instructor Course. At the heart of an effective training pace is the accurate selection of scenarios and vignettes that are predictable and measurable in mental and physical task inventories.

Current efforts at the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Division (AFRL/HEA) are transforming the art of scenario selection into objective principles through definition of mission essential competencies (MEC) (Colegrove & Alliger, 2002). Continuing research in defining MEC knowledge, skill, and experience (KSE) inventories enables warfighting trainers to dissect scenarios into decision-task sequences (Bennett & Crane, 2002). Each sequence can then be viewed in the context of a scenario to define required elements and elements subject to modification or elimination in a calculated risk of priorities during combat operations. The decision-task sequence a warfighter chooses to employ in a scenario is observable and predictable when assessing weapons employment in terms of objective standards and MECs. Each task sequence is a finite set of decisions and actions that occur in sequential and parallel execution during employment. These in turn can be mathematically modeled into a relative complexity value using an anchor point of the most elementary actions required in a tactical situation. Given this emerging resolution of complexity, the warfighter and training program developer will be able to accurately scale the complexity in training scenario vignettes used in a structured program.

Research Benchmark

Previous AFRL research into complexity modeling (Crane et al., 2001) established a first look at quantifying complexity by using a variety of operators and levels of experience to stratify a select number of scenarios utilized in DMT. An empirical method was chosen over analytical because the analytical approach was insufficiently definable among subject-matter experts (SMEs). Empirical rank orders of each group demonstrated substantial agreement among evaluations with correlation values from +0.96 to +0.97 between less experienced pilots, instructor pilots, and air weapons controllers. In assigning a normalized complexity value to the rank order and applying it to the existing, and very low resolution grading standard, trainee progress in a building-block syllabus of instruction was quantified. The empirical method demonstrated that a complexity factor would aid training progress visibility, but not without some questions as to its origination, method, and scale anchors. Normalized rank ordering of selected scenarios limits the value definition to the selected set. New scenarios would have to be shuffled into the deck and the deck reassessed by more input from SMEs with the new assessment being subject to personal perspective of new SMEs.

Individual pilot perspectives on training and combat experience has, to this point, made training and execution methodology consensus among operators difficult to achieve. This, in turn, makes definition of complexity difficult for researchers to grasp when subjected to operator disagreement or lack of instructional and employment familiarity. In spite of Crane et al.'s concern regarding previous applications of analytical methods for complexity scoring, a more rigorous analytical approach may generate reliable and valid complexity scores while avoiding the difficulties of empirical scaling. The empirical method is cumbersome to use, sensitive to the pool of scenarios and SMEs selected, and can produce different magnitudes and scalars.

In the last year, key breakthroughs in the areas of MEC definition (Colegrove & Alliger, 2002) together with employment and training standardization within the F-15 and F-16 communities have opened routes for more detailed complexity analysis. Crane et al.'s (2001) critical assessment of analytical scaling was carefully considered during the development and

evaluation of this approach. The following sections describe the assumptions and logic for the method of complexity analysis and validation discussed in this presentation.

METHOD

An analytically-based complexity estimator was derived after study of the elements of a tactical problem using the MEC construct to examine actions, decisions, and communications involved in tactical situations. The resultant model was built with the intent to accurately capture performance standards in USAF F-16 training programs. The analytical model aligns to the common training methodology used in fighter programs. Before detailing the model, it is essential to understand the assumptions resident in “realistic” training.

The tactical objective of an *actual combat mission* resides within the commander’s intent as stated by an air tasking order. The objective may not call for complete elimination of adversary forces (although that is commonly fighter pilot’s coveted end state). Typically, the objective of real-world missions involving air combat will be to establish air superiority, block incursions into an operating area, and deny the use of selected airspace for some period of time. The degree of aggressiveness of adversaries, combined with logistical constraints of weapons and fuel, and psychological pressures on both sides will temper the degree of confrontation. In actual combat operations, there may be less killing of targets and more chasing away and monitoring of the area. While this is often acceptable for the real combat objective, training to these outcomes is less desirable than pushing to the limits of execution capability.

In “realistic” training, the objective of a scenario is typically raised to the most stringent execution standard of eliminating all incursions into an area of responsibility. Additionally, adversary conditions are set to predetermined levels of offensive pressure, defensive reactivity, and shootable merit (as measured by rules of engagement) to challenge and improve the trainee’s abilities to operate effectively. This training methodology is hallowed practice of air combat training in all US fighter forces (USAF, USN, and USMC) and is hence the foundational assumption of establishing a tactical objective in complexity definition – the team’s job in training is to detect and kill all adversaries coming into the area of responsibility, and they *will* come by scenario definition. Distributed mission training scenarios at AFRL are built upon this customary practice. The assessment of complexity assumes that each adversary group must be assimilated into the tactical plan and dealt with by the fighter team at a time deemed appropriate for the tactics employed on both sides.

Complexity Value Derivation

Physical and mental tasking studies within AFRL and other sources were reviewed in concert with compilations of instructor commentary and training reports from the US Air Force Weapons School (USAFWS) F-16 division over a three-year period. This time span involved six Weapons Instructor Course (WIC) classes, 68 students, and 35 instructor pilots. An additional two-year period was assessed in basic F-16 qualification courses from instructional and training development documentation. Analysis consisted of examining commentary and compiling consensus on attributes of acceptable versus non-acceptable performance based on current F-16 USAFWS tactical employment standards. These attributes were then compared to the KSEs

within each MEC to construct basic decision-task sequences or KSE execution paths required to accomplish a tactical objective for each of a finite set of adversary positions, characteristics, and maneuvers. The number of execution paths chosen follows the *element fire and maneuver doctrine* outlined in the F-16 employment standards. Within each execution path, the decisions and tasks were analyzed for sequential and/or parallel characteristics along with an expert analysis of automaticity versus the need for conscious problem solving priority. Models for elemental attributes within a scenario were then constructed using a baseline value for the simplest or most automated decisions and tasks using a mathematical construct to combine the elements into a baseline score for the simplest execution paths. Several differential approaches were considered; however, the situations in air combat do not lend themselves easily to differential analysis with the limited data sets available. Instead, weighted values, combined in products, summations, and limited exponential modeling proved to be a simple and reliable method to express the scale factors sought.

Baseline complexity for the simplest intercept type was molded from a consensus of F-16 WIC instructor considerations defining the most rudimentary tasks and pacing found in an air intercept situation. The baseline attack may take the form of either a simple straight line closure to a weapons envelope, or include a maneuver called the *baseline intercept* to a stern conversion. Air intercepts in distributed mission training and actual combat can rapidly take either shape due to weapons limitations or scenario rules of engagement. In forming the baseline consideration, the lowest task count, load, and pace was considered in the context of the skill sets of both types of intercept to establish the lowest level of complexity to achieve a kill objective.

The main criteria in evaluating both intercept types was to define a point from which the tactical experts train forces to employ weapons with the highest probability of weapon effect and the most automaticity in execution. For beyond-visual-range engagements, this position is a collision course to the target followed by a range preserving maneuver. For the visual identification intercepts, expert opinions varied from ending in a position abeam the target to ending in a position aft of the target. However, the baseline intercept remained consistent in skill set and pace for both end states prior to weapons employment in quantitative terms.

A foundation score of 1.0 for an intercept of least complex execution was selected to facilitate calculation of more difficult tasks. The baseline intercept and collision closure were assessed to have comparable task quality and quantity since both required detecting a target, enabling a radar track file through switch activations, and continuously analyzing stable mathematical relationships displayed on the air-to-air radar's B-scan display to adjust for and maintain the desired intercept profile. In the baseline score, assumptions are made for an easily detected and discerned target, no counter-maneuver or awareness of the fighter by the target, no weapon employment restrictions, and no detractors to situation awareness development. The foundation score represents the simplest task set an individual fighter pilot will encounter in a tactical scenario while developing minimum required situation awareness and physical positioning to achieve a kill. Both versions achieve high automaticity levels in early tactical training.

Scenario Complexity Estimate

Models of higher order complexity values for adversary attributes beyond the baseline were initially estimated through analysis of select execution paths used to establish the USAFWS F-16 Division employment standards. These execution paths encompass the realm of options an *element* (two aircraft operating as a team) would choose from when confronted with varying levels of adversary capability and maneuver. An initial value set was determined by analyzing observed performance in video-taped tactical displays and Nellis Air Combat Training System (NACTS) data collection. Determination included extensive analysis within the expert instructor's debrief of performance evaluation along with assessment of decision and task execution quality as recounted in the mission summary and grading documentation. Weighting factors considered additional mental and physical tasking of individuals within the *element* firepower team, essential communications requirements between team members, and limitations imposed on available resources posed by additional stressors in or near the fight space. Each execution path was then calculated by the model and submitted for independent review by SMEs at the F-16 WIC. As a result, initial estimates of certain weighting factors were reduced in value, some stressors fractured for better resolution, and several were increased in weight value. Attributes assigned to single groups of adversary aircraft include separate weighting factors for number of aircraft, detection range, sort ability at range, weapons carried, awareness of fighter attack, counter-maneuver, countermeasures, and shootable merit based on rules of engagement. These factors were derived from the F-16 tactical standards' cubic construct of awareness, reactivity, and identification as measures of stress imposed by the adversary force on a fighter team's tactical problem. Since the single-group construct followed standardized employment methodology, consensus for single group complexity valuation proved consistently high in SME review.

Estimation of complexity in the empirical method was highly sensitive to the total number of groups in a scenario (Crane et al., 2001). Crane et al. suggest that group number may or may not have a standardized complexity stress; however, the empirical data analysis suggests the experts have significant agreement in estimating group-to-group complexity by numbers alone. The empirical rank order of scenarios was almost completely stratified by total group number with most disagreement in stratification falling primarily in same-number scenarios. In tying groups together in the analytical method, consideration is given to the distance between groups using the same decision process as the engaging pilot. Range sensitivity is expressed by the combatant in terms of delineation points known as *bounding range* and *factor range*. *Bounding range* defines the decision to include or exclude a group from the tactical situation. *Factor range* defines a break point in attention paid to a group when dealing with a primary target or threat. Groups inside *factor range* are given higher complexity weighting as a function of their proximity to an *element's* primary tactical problem. A *factor group* is part of the tactical problem even if engaged by another force since a *factor group* represents a pending threat that *must* be considered. The analytical segment built for group-to-group interaction provides a straightforward accounting of group positions and ranges for consistent valuation. The construct chosen retains reasonable sensitivity to total numbers of groups, but simultaneously allows for accounting and rational limitation of complexity inflation for groups that are not part of the immediate tactical problem.

Large scenarios (more than four groups) remain problematic in assigning complexity to distant groups being engaged by distant friendly forces and sequential attacks by waves of adversary forces. Initial input from subject matter experts suggested a summation strategy was appropriate. Statistical analysis of this method proved it less appealing than a weighted strategy. The current methodology for scenario analysis is an exponential weighting approach based on MEC KSEs and standardized employment. Standards and MECs were used to resolve an inventory of all tasks, decisions, and inter-flight communications for commonly executed adversary presentations. A procedure for generating complexity estimates for multiple groups was selected that retained the value of attributes of individual groups and assigned a limited exponential curve to increasing numbers of groups. This procedure provided optimum retention of correlation with expert opinion while avoiding statistically troubling characteristics.

ANALYSIS

Trials conducted on the complexity estimate model began with a comparative analysis with the Crane et al. (2001) empirical research, then proceeded to a validation test with new AFRL benchmark training scenarios, and continues with analysis of the benchmark scenarios in DMT along with an analysis of live-fly and simulation scenarios resident in the highly structured training program of the USAFWS F-16 Division air-to-air training program. DMT syllabus building efforts are also using the complexity estimate as additional guidelines for scenario sequencing.

Empirical Comparison

Complexity estimates were conducted on the same scenario set as presented in the empirical study. The estimates were produced in a blind study without prior knowledge of the empirical stratification. Analytical results were then compared to the empirical study results. A strong positive of correlation (+0.96) was achieved with the complexity estimates derived from subject matter expert rankings in the empirical study. All scenarios except three remained within one standard deviation of the previous study. Complexity index values were allowed to expand as the model dictated and achieved values from 1.00 to 54.00. Modeling (M1) also initially indicated a capability to visualize training plateaus suggested in Crane et al.'s empirical research. The values are shown in Figure 1.

Crane et al.'s empirical procedure stratified scenarios in a subjective valuation of the entire scenario as viewed by each study participant. The initial analytical estimates were also computed in a holistic sense using the entirety of scenario attributes in a combined score. As the scenario expanded in size and numbers of stressors, the complexity value inflated exponentially. While this may seem intuitive to the analyst, SMEs disagreed with the valuation scheme. The SMEs indicated that the entire scenario plays a lesser role in tactical problem complexity than Method 1 analysis delivered.

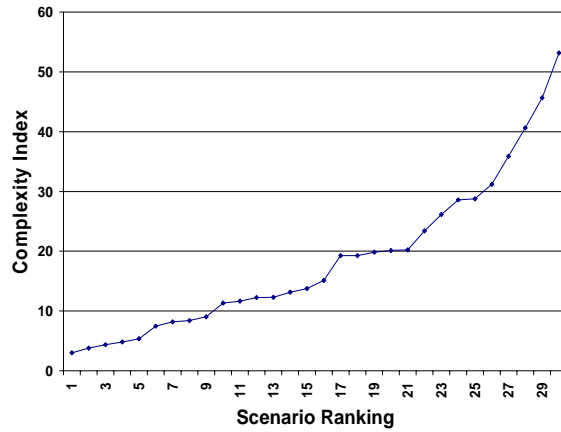


Figure 1. M1 Analytical Complexity Value by Scenario

Additional scrutiny of fighter air-to-air employment methodology suggested that the model needed to be able to break complexity elements into aggregates as the fighter pilot would detect and deal with them. The complexity model has sufficient flexibility to segregate and value engagement vignettes as well as total complexity estimates. A second methodology was examined and recalculated using the subject matter expert inputs. Method 2 (M2) examines the scenario from the weighted approach. Comparisons of empirical and analytical stratification for Method 2 are shown in Figures 2 and 3. Figure 2 shows a summation approach to total scenario complexity while Figure 3 shows the weighted approach.

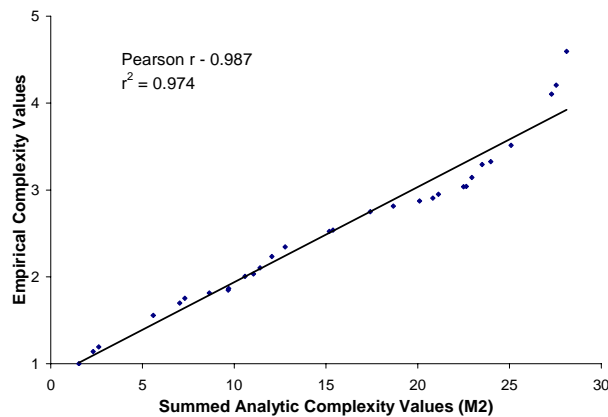


Figure 2. Comparison of M2 Summation Complexity Estimates with Empirically Derived Estimates

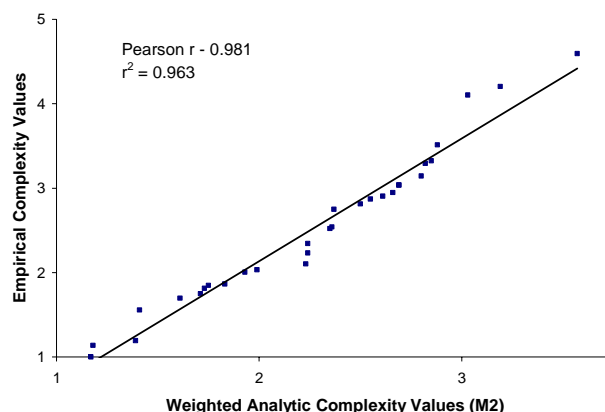


Figure 3. Comparison of M2 Weighted Complexity Estimates with Empirically Derived Estimates

The results using summation in Method 2 suggest additional investigation is required into the non-linear relationship seen in the comparative graph. The M2 weighted method shows a more random distribution around the regression line with little change in correlation with empirical complexity scores. These analyses suggest the M2 Weighted procedure is a more promising avenue toward complexity factor valuation.

Benchmark Reliability

A second blind study was initiated by DMT researchers to evaluate the ability of the model to reliably compare training scenarios of equal complexity. The researchers provided ten benchmark training scenarios developed by SMEs and the AFRL DMT staff to evaluate pilots under the current DMT research program. These scenarios had been scrutinized and flown in DMT to validate the outcomes and pilot perspectives in the scenarios. The ten benchmarks were actually five parallel or mirror image scenarios that were properly identified by the model. Thumbnail sketches of selected scenarios are shown in Figure 4.

Calculation of the benchmarks resulted in summed values ranging from 9.04 to 9.40 and weighted values from 2.77 to 2.88. The values for composite scenario complexity are shown in Table 1.

In addition to demonstrating reliability in aggregate calculation, this research also tested the ability to extract complexity scores for individual engagements by element firepower teams. Using the M2 estimate to view and calculate tactical problems as the fighter team sees them, and applying the same calculated risk of disregarding certain *non-factor* geographic attributes as would be done in the immediate situation by the fighter team; the M2 estimate scores reflect the same complexities of smaller scenarios that were considered accurate in previous subject matter expert reviews. It is essential to understand that the surveillance and targeting risks taken by a fighter team in a geographic sense are attributes found within an attack wave, but either handled by another element team, or outside of factor range to the element team being evaluated.

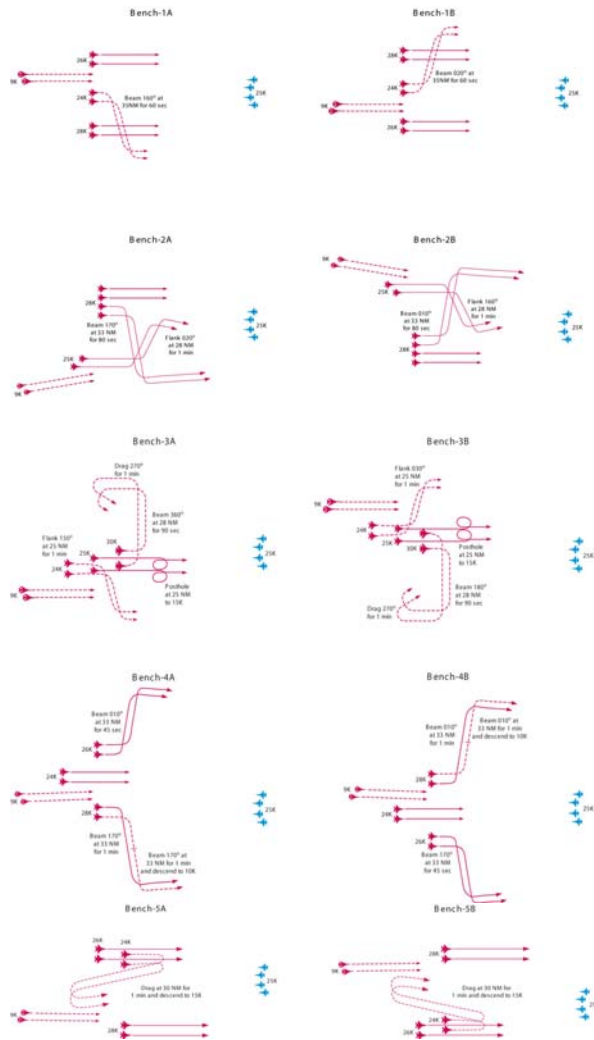


Figure 4. AFRL DMT Benchmark Scenarios 1-5

Scenario #	Sum Value	Wt Value
1A and 1B	9.04	2.77
2A and 2B	9.15	2.80
3A and 3B	9.40	2.88
4A and 4B	9.40	2.88
5A and 5B	9.06	2.77

Table 1. Summed and Weighted M2 Scenario Complexity Values for Ten Benchmark DMT Scenarios.

Another segregation attribute in fighter employment was modeled in the M2 benchmark test based on *non-factor* attributes separated by time. Fighter teams will take on a workload in each engagement that balances situation awareness, optimized firepower, and defensive status. As the team reaches culmination in an engagement, the leader will direct a short retirement

known as “*cold ops*” to separate the fighters from threats and allow time to rebuild range and/or situation awareness for reattack. Similarly, certain tactical situations influence fighters to plan sequential attacks from a maneuver sequence called “*the grind*.” The model was used in both geographic and time divisions to assess lower levels of complexity internal to the complete scenario.

Separated scores demonstrate not only *element* complexities, but also suggest the model can be used to describe complexity per attack wave as determined by timing and maneuver of each side’s tactical plan. In the case of the benchmark scenarios, *element teams* would be compelled to simultaneously attack in azimuth, followed by a short retirement and reattack of remaining adversary forces. The tactical method chosen to break down the internal complexities is a direct reflection of current F-16 WIC school solutions and employment guidelines. The results of the benchmark analysis computed by the complexity estimate model M2 are shown in Table 2.

Scenario #	N Fighters	S Fighters	Reattack
1A	2.24	2.70	1.76
2A	2.45	2.70	1.76
3A	2.70	2.57	1.76
4A	2.57	2.71	1.76
5A	2.71	2.24	1.76

Table 2. M2 Benchmark Complexity by Fire Team

Weapons Instructor Course Analysis

Scenarios provided from the empirical study were of insufficient detail to accurately test the full attributes of any of the analytic complexity estimator versions. Moreover, the narrow scope of the attributes applied in the empirical method substantially restricted the range of application for, and the long term utility of, the research data. As a result, critical details for estimation were subject to changes in the perspective of the original group of empirical study participants and were considered a possible cause of some of the larger sigma values in Crane et al’s (2001) study. An additional concern is that the DMT threat model used to create training scenarios is limited in comparison to live-fly targets. Live-fly threat pilots are briefed to execute precise tactical maneuvers that the threat model cannot duplicate at the present time. Focused training at the F-16 WIC relies heavily on establishing tactical vignettes with precise attributes and considerable threat validity and realism. Any complexity scoring approach must be able to fully account for a full suite of threat and adversary tactics if it is to be useful for valid virtual and live tactical training.

The F-16 WIC formalized a standard approach to adversary presentations beginning in 1988. Since that time, the practice has grown to include all aspects of air-to-air training within the syllabus. Phase managers provide all participants with exact presentation attributes requested for the WIC’s structured ascent to graduate-level proficiency. However, the selection of scenarios and attributes in WIC training still relies exclusively on historical precedents of

scenarios chosen in a subjective manner and never analyzed in any other method than a brute force of trial and error.

The next evolution of testing for the complexity estimator was to subject its analytical technique to a sample of air-to-air instruction scenarios existing in the F-16 WIC. The sources for these scenarios are the *bandit playbooks* from *Tactical Intercept* training, academically presented scenarios used to teach air combat concepts, as well as graded observations and playbacks of Nellis Air Combat Training System recorded missions from the more loosely structured *Air Combat Tactics*, *Weapons*, and *Mission Employment* phases of instruction.

Twenty scenarios were selected from WIC sources including one well-recognized baseline problem from initial qualification training. The 20 scenarios were analyzed under the M2 technique using lessons from the empirical comparison and DMT benchmark studies. The goal of this evaluation phase was to compare live-fly complexity values to the consensus opinion of F-16 WIC instructor pilots and training phase managers based on where these scenarios appear in the progressive training of the WIC. It should be noted that the sample size of 20 scenarios is considered too small to draw firm conclusions at this time and additional research using a larger sample will be undertaken to investigate the preliminary outcome of this analysis.

Results of the WIC sample analysis determined M2 weighted complexity values ranging from 1.01 to 2.91. Using the M2 summation technique, the scores ranged from 1.01 to 9.51. Comparison of weight to summation was analyzed to identify how numbers of groups affected the overall rank ordering from technique to technique. The results are shown in Figure 5.

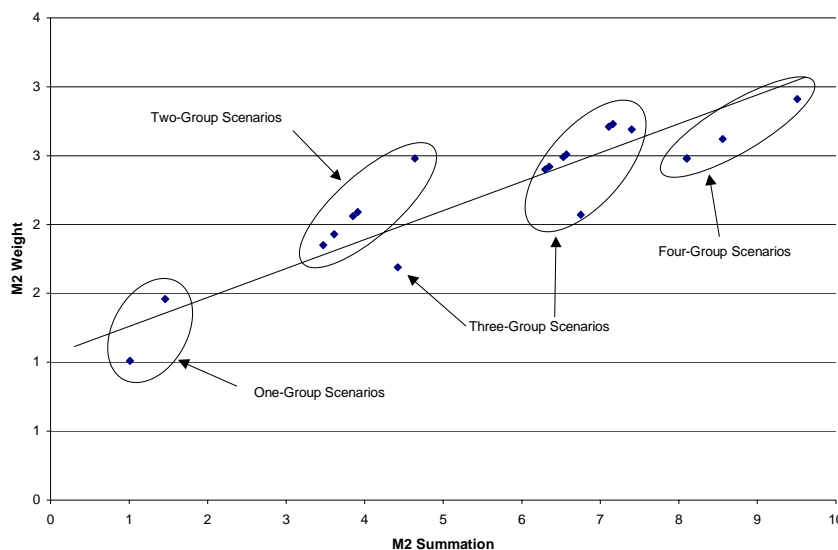


Figure 5. F-16 WIC Scenarios – Complexity Technique Comparison of Weighted Groups versus Summed Groups

The next phase of analysis was to examine the complexity scores in the weighted method in comparison to the sequence they may appear in the WIC syllabus of instruction. F-16 WIC air-

to-air training is broken into segments of multiple rides that examine and train weapons officer candidates in specific employment areas. Training phases are engineered to have increasing levels of stress that may be raised from scenario to scenario or mission to mission to enhance or test the learning pace. The *Tactical Intercept* phase (TI 1-3) is a three-mission block of instruction designed (due to logistics constraints) to ramp students quickly up to radar and weapons proficiency. The first half of the *Air Combat Tactics* phase (ACT 1-3) is the first opportunity to put an entire mission plan to practical application and is hence flown against adversaries with moderately complex attributes. The second half of the *Air Combat Tactics* phase (ACT 4-6) introduces students to the most robust adversary weapons and tactics attributes. The *Weapons* phase copies the air combat problems of the *Air Combat Tactics* phase while introducing the multirole aspects of F-16 employment. The final phase of the course, *Mission Employment* (ME 1-3), provides the largest and most complex problems for students through addition of the requirement to coordinate many friendly platforms other than F-16s into the scenario. Since the complexity estimation technique is yet untested in multi-mission environments, only the air-to-air problems faced by F-16s were extracted from the final two phases for analysis.

Complexity scores for the 20 scenarios were assembled and linked to the missions they relate to either through a bandit playbook or actual observation. Results of a weighted comparison using M2 are shown in Table 3.

Scenario	M2 Weight	M2 Sum	Mission Training Block Linkage				
6	1.01	1.01	IQT				
1	1.46	1.46		TI-1			
5	1.69	4.42		TI-2/3			
24	1.85	3.47		TI-2/3			
2	1.93	3.61		TI-2/3			
18	2.06	3.85			ACT 1-3		
13	2.07	6.75				ACT 4-6	
20	2.09	3.91			ACT 1-3		
10	2.40	6.30				ACT 4-6	
16	2.42	6.35			ACT 1-3		
12	2.48	8.10				ACT 4-6	
14	2.48	8.10					ME 1-3
19	2.48	4.64				ACT 4-6	
11	2.49	6.53					ME 1-3
8	2.51	6.57					ME 1-3
7	2.62	8.56					ME 1-3
9	2.69	7.40				ACT 4-6	
17	2.71	7.11				ACT 4-6	
3	2.73	7.16					ME 1-3
15	2.91	9.51					ME 1-3

Table 3. F-16 WIC Scenario Mission Linkage by Training Block.

The F-16 WIC analysis suggests that the complexity estimation technique follows subject-matter expertise used by the USAFWS to structure the training program for WIC students. The selected scenarios demonstrate attributes that generally follow the easy-to-harder goal of each training segment with some new outlooks on training pace. A time-honored tradition of allowing the adversary force substantial latitude in scenario selection and execution within a three-mission training segment inserts a considerable variance in scenario pacing within

the F-16's ACT, WPN, and ME phases. The limited sample size for this investigation also serves as a warning that more detailed investigation is required to produce a high confidence level in the estimation technique. At the same time, the analysis also suggests there is reason to reconsider the trial-and-error method as well as adversary latitude of scenario selection and opt for combined techniques using both subject matter expertise and analytical complexity estimation in the selection of scenarios and vignettes for tactical intercept and air combat training.

The next phase of research will be to collect a larger group of WIC scenarios utilizing all bandit playbook options from the *Tactical Intercept* phase and as many observed adversary presentations as possible from the *Air Combat Tactics, Weapons, and Mission Employment* phases. This research is in progress and will be fully explored as a goal to establish the full level of complexity definition and analysis in F-16 air-to-air employment training.

Applicability to Other Aircraft and Missions

Another question being considered in complexity estimation analysis is the applicability of the method to other aircraft executing air-to-air combat and also to F-16 missions that include objectives in addition to or exclusive of air-to-air. Preliminary review by F-15 tactical experts suggests there is differing opinion on complexity valuation in the model. Work in this area requires a complete understanding of the perspective used to create the estimation values from each aircraft's pool of expertise as well as applied analysis similar to the F-16 efforts so far. Researchers at AFRL are preparing for studies in F-15 applicability and valuation as well as F-16 multirole missions that include surface threats, air-to-surface elements, and environmental stressors such as electronic warfare and systems attrition through failure or combat loss.

The physical considerations that drive a high expectation of broader applicability is the strong similarity of current fighter systems and the common set of weapons shared among them. There is also high expectation from an institutional standpoint. The tactical employment standardization of F-16 forces followed from studies and alignment of tactics recommended by the F-15 WIC and tactics manual and the F-18 TOP GUN instructor staff (and corresponding TOP GUN tactics manual). Therefore, along with common systems, all three aircraft enjoy a relatively common employment doctrine and methodology in air combat. The implications of this common doctrine and methodology are considerable generalizability and leveraging of the methods and tools developed in this complexity effort for other domains of similar complexity.

Broadened application to other roles and missions requires investigation of MECs and KSEs for those areas and the establishment of doctrine-based decision-task sequences required to successfully achieve the objectives of each area. Definition of decision task sequence paths is the key to establishing consistently expected and measurable levels of complexity. The focus of complexity research will remain on the F-16 until achieving credible confidence levels for the basic model; however, probes into F-15 employment and F-16 multirole missions are opening the gate to larger fields of application.

Lessons for Future Employment

An enduring issue in developing tactical complexity estimates is establishment of a framework for how researchers must approach the basic character of combat tactics and the operators that conduct them. Two main themes emerge in considering complexity employment in DMT operations: (a) to expand an objective measure of complexity into other mission areas or multirole operations, it is necessary to examine and codify the discoveries of this effort that define both its successful attributes as well as limitations; and (b) to use complexity valuation as a tool for cataloguing scenarios for training intensity, complexity estimates must achieve as close as possible to realized complexity as experienced by operators training in DMT.

The previous section noted potential for applicability to other similar airframes. The belief in this ability to cross platforms is rooted primarily in the high level of commonality of basic method in each community. When expanding to other missions or roles, complexity studies will be required not only to define what they can do but also where the technique is likely to lose its validity and reliability. Likewise, this journey into complexity valuation observed variations arising from employment of diverse tactics and techniques within the same scenario. As research progresses, it is becoming more clear that two complexity values will probably emerge. The first is a static average or “off-the-shelf” value for cataloguing available training stresses (the primary work of this research), and the second is a real-time calculation for individual and team performance that accurately portrays complexity as delivered by choices made by the team during scenario play-out.

One of the first obstacles to establishing valid observations was clearing up conceptual fog regarding *difficult* versus *complex*. Observing in other disciplines that *complex* items are usually not *difficult* for a master, while *simple* tasks may prove *difficult* to a novice, it was clear that objective complexity needed to be employed to eliminate sensitivity to *proficiency* and/or *automaticity* if these are indeed the targets to be evaluated in training. Operators providing subject-matter expertise for this research did not, at first, have a clear understanding of the difference between complexity and difficulty. This led to some reorganization of the data collection effort in early stages to reduce variance of subjective observations.

The objective of establishing *complexity* measures is to create a valid measure of actual task and decision inventories for a given situation. *Difficulty* resides more in the cognitive domain as a measure of how an operator’s proficiency in an event measures up in handling the realized complexity of the event. While conducting the initial observations of operators in a given task sequence, the subjective grading of complexity as the operator saw it, generally declined in the range of 5-10% as that operator’s skill proficiency and foreknowledge gained through repeated exposures to the same event. Events scored at higher values of complexity at first attempt were often devalued by the same operators in debriefs of later trials of the same situations. The range of deflation peaked mostly in situations that were not normally trained in live-fly events, but were tested repeatedly in Unit Training Device (UTD)¹ trials to establish baseline values of scenario attributes. Task sets seen more regularly in live-fly tended to be reduced to a lesser extent.

¹ Unit Training Device – stand alone simulation device used for tactical training in fighter units (the predecessor to DMT devices).

A second set of operators was employed to investigate complexity scoring after providing a detailed discussion of objective complexity and how a pilot's proficiency gains may tend to make the tasks easier. The research observer instructed the operators in the framework of mission essential competencies and individual task/decision sets to establish a framework for thought in determining what the pilot *actually had to do* to achieve the desired outcome for a given event. In each complexity debrief, the second set of operators were asked to examine the MEC task/decision list and provide the sequence they employed to meet the objective of an observed event. In the second set of operators observed, complexity evaluations showed a higher level of consistency when starting the individual- and team-level task definition lists.

Examination of team-level interactions employed by operators in distributed training highlighted as unattainable the baseline goal of complexity scoring as stated in the initial concept task order. The original intent was to provide a preset value of complexity that would extend to all activities within the scenario. Observations at the team level during validation testing (leading to the M2 method) showed wide variations in actual complexity faced by components of the four-aircraft teams depending on flight leadership, tactics selection, adversary force posture, and weapons effects. Establishment of a static complexity value for a tactical scenario suggested overly broad reach for validity and reliability given the wide variance of possible complexities over time and well as due to the choices made by the leader and other teammates. The main objective of complexity definition, a way to discriminate actual performance levels, caused the research effort to alternately study a method that follows the geography of increases and decreases in complexity during the time span of a tactical scenario.

To shed more light on the need for real-time complexity valuation, consider the evaluation of proficiency of an element (3,4) of a team of operators who, through a decision of the flight leader, is retired to a grind tactic in the M2 benchmark scenario 5A. The element is taken out of the first attack wave and will be the sole operators involved in the second attack wave. The situation developed as a result of the flight lead's decision demonstrates the M2 technique's sensitivity to tactical execution. Rather than the complexity values as shown in the earlier section where all four team members align on the attack axis each cycle, the initial attack is solely conducted by the flight lead and his wingman and is elevated to much higher short-term complexity both as individuals and an element fire team. The following tables and figure highlight changes found in the simple decision to switch the attack plan to one different from the *recommended* solution.

Scenario 5A	N Fighters	S Fighters	Reattack
Team CI	2.71	2.24	1.76
Ind. CI (L/W)	2.24/2.86	2.24/2.24	n/a

Table 4. M2 Benchmark Complexity by Fire Team

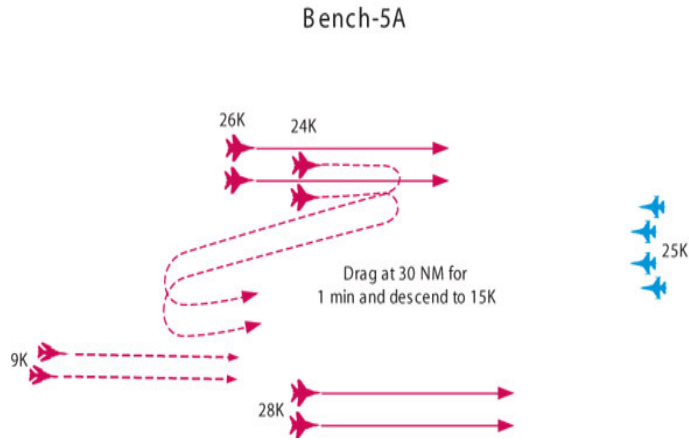


Figure 6. M2 Benchmark Scenario 5A

The changes in complexity resulting from a change in attack tactics result primarily from the loss of fighter resources to share the load. The adversary situation remains the same and the remaining fighters must incorporate higher numbers of adversaries per fighter into their attack plan.

Assessing results of the first wave using the most probable outcomes as seen in research evaluations and operational training missions, it is likely that the number of adversaries will remain high for the second attack wave. For classification reasons, the issues and impacts will not be fully resolved and described here. The resulting complexities established in real-time for the decision made is notably higher as shown in Table 5. These complexities will be owned completely by the fighter noted as opposed to sharing the complexity of execution among a wider force.

Attack Wave 1	N Fighter (2)	S Fighter (1)
Real-time CI	2.98	2.46

Table 5. M2 Benchmark 5A First Wave by Element

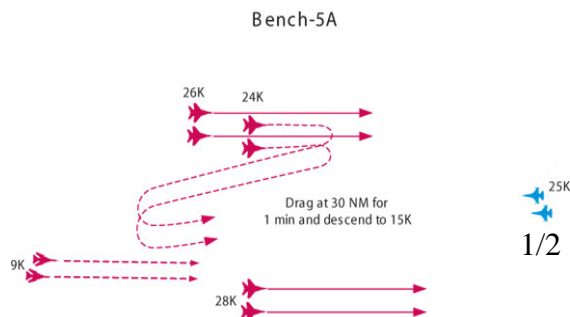


Figure 7. Bench 5A Second Wave

Attack Wave 2	N Fighter (4)	S Fighter (3)
Real-time CI	3.25	3.03

Table 6. M2 Benchmark 5A First Wave by Element

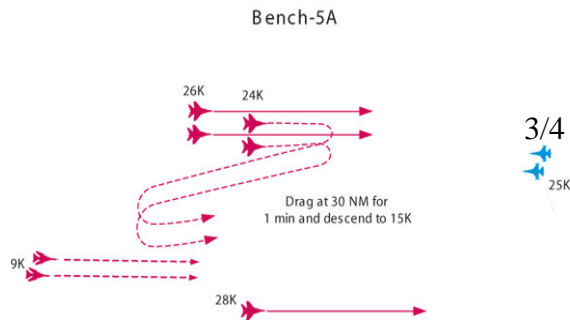


Figure 8. Bench 5A Second Wave by Element

Applying the probable outcome of the first wave attack, the projection for the next wave conducted by #3 and #4 in the team is considerably more complex and would require a much higher level of proficiency to reduce the numbers. In the first tactic, the fighter team employs as a group to reduce numbers to a lower complexity for the subsequent attack wave. In the second example, a piecemeal attack on the same picture begets vastly higher complexity for the second element's attack due to the inability of the first attack to reduce adversary numbers in the same manner as the recommended solution.

While these examples demonstrate sensitivity of tactic selection to the level of complexity realized, the message is often not as clear to the operators whom we engage as research subjects. In spite of recommendations and guidelines delineated in tactics references and taught at the weapons instructor course, other political and psychological forces are consistently noted by DMT researchers at AFRL/HEA when observing flight planning and operations. When establishing a framework for how to define complexity, the outer layers of apparent dissention and disagreement among operators must be penetrated to achieve valid and reliable results. The roadmap for objective determination was developed in parallel from the decomposition of mission essential competencies for air superiority. This in-depth extension of the work produced by Colegrove & Alliger (2002) proved a cornerstone effort in resolving apparent conflicts arising from divergent operator viewpoints.

Tactic selection is not the only sensitivity to consider in complexity. A similar result can occur from reduced weapons effects even in the recommended tactic. In the event shots taken do not produce the projected attrition, complexity definition cannot be considered reliable for the reattack when the actual scenario in the second attack varies moderately from the planned reattack picture. A fact of life in tactics is the ever-present enemy desire to cause a plan to fail. These forces create all manner of inefficiency of targeting and weapons employment. An accurate complexity estimate must take into consideration all that influences the attacking flight

in each wave. In the projected complexity for follow-on attack waves so far, certain assumptions have been made about weapons effects. The projected killed adversaries do not weigh the second attack's complexity estimate. When assumed kills do not materialize, the picture is more complex in real-time and must be analyzed veritably to assess the actual level of performance exhibited by the operators.

Another situation to consider involves an imbalance in a fire team beyond the normal expectation of the tactics manual. Fire plans are developed with consideration for the proficiency levels of *average* operators. The average operator as defined by the USAFWS units involved in air superiority training models the *average operational team* of shooters as a leader with solid capability and a wingman with moderate skill development. In practice, these types of teams are accompanied by many other combinations of leader and wingman experience levels in DMT research. Conventional wisdom practiced by the USAF in assigning duties to the individual members of the team follows a model taught by the USAF Flight Safety Center and best elaborated by Edwards (1975). Edwards' concept of proper *trans-cockpit authority gradient* extends to fighter elements where the element lead may be considered the higher authority in decision making and at a more advanced stage of knowledge and skill development, while the wingman is considered to be less adept at decision making and skill proficiency. This modeling of leader and semi-skilled follower is a primary assumption of all tactical methods and plans. However, contingency tactical plans often resort to battlefield initiative of wingmen who fall outside the normal boundaries of skill development. In many cases, both planned contingency branches and observed DMT research scenarios demonstrate duty crossovers which affect not only the outcome of the mission, but also call into question how the team should be assessed in terms of standard leadership authority and duty allocation.

At issue in valuation of a non-standard team protocol is the ability to capture real-time complexity values in the condition of an individual obviating another team member's need for action by taking on a higher level of tasking individually in certain contingency operations. This behavior can skew realized value away from projected results and reduce the validity of "shelf" estimates made before the event. An invalid estimate then runs the risk of improperly calculating the performance measures sought for individual skill assessment. While it is considered in the current model that a flight leader may backup wingman shortcomings, the complexity estimate must also be ready to consider the opposite case as well as more extreme situations than the assumed tactic may call for. These occur as a result of execution inefficiencies and regularly enough in tactical operations to warrant a place in the definition process. A static complexity estimate may average the team score in this case rather than elevating the *save* conducted by one member to its rightful higher level of individual performance. At the same time, poor performance of an individual should be properly devalued when observing team contingency actions. A static number covering the entire scenario is unlikely to be able to produce either value reliably.

A promising answer to real-time complexity valuation is emerging through development of an *expert* human performance model. While many of the parameters needed to define attributes in the complexity estimator exist as basic data in DMT architecture, how the flight cues on data and decides to attack is neither resident nor obvious. The HP expert may allow DMT developers to incorporate real-time complexity valuation into performance assessment. Trainers

will still be confronted with a deviation between expected and actual performance, however, with real-time estimates, the ability to show critical execution paths in a *school solution* versus actual complexity may prove to be a valuable tool for developing team leadership as well as peaking individual performance. Study continues on the time and scenario segmenting strategies that will prove most reliable for achieving optimum valuation.

CONCLUSIONS

The drive for increased efficiency in tactical training demands that training developers and instructors on the front line understand the role of complexity in training pace. Moreover, accurate application of complexity valuation must also be defined and applied if its inclusion to performance grading standards is to accomplish its intended goal. Crane et al's (2001) study highlighted the inability of subject matter experts to agree on valuation and combination of complexity into a single score for a scenario. This observation was made before MEC KSE linkage was fully developed in the light of decision-task sequences as laid out by the expertise in the USAF F-16 tactical employment standards. Given the advances in understanding execution elements resultant from MEC analysis, subject matter experts are now closing on an ability to not only stratify scenarios by complexity index, but also to provide a qualitative valuation without need for repeated empirical trials. The lessons learned in this ground-breaking effort also provide valuable insight on the need to develop real time complexity indices for all levels of team and individual performance as well as the projected estimates to catalog training scenarios for call up in DMT training in the future.

Along with the ability to remove floating values of complexity resident in empirical methodology, the M2 method also shows promise for internal analysis of selected engagement vignettes to adjust training paces for individual participants within a larger scenario. In other words, we can adjust the complexity and pacing of scenario events and ensure that the training delivered in DMT environments is adaptive to *each* learner's needs and to the learning needs of the smaller team. There are two key goals of complexity analysis. First is to enable training program developers to accurately adjust learning pace in resource-constrained ISD environments. Complexity analysis linked to demonstrated performance will provide substantive data to make the case for enlarging or reducing training resources to meet specified objectives required of the gaining combatant commands. The second goal is to enable the front-line instructor charged with executing DMT training to visualize and optimize progress through learning cycles and plateaus by making real-time changes to training scenarios. That is, it will soon be possible for researchers and instructor pilots to accurately diagnose knowledge, skill, and experience "gaps" and to identify specific scenarios in both virtual and live-fly training to target and eliminate the gaps. By recognizing and systematically emphasizing proper execution in weak performance areas, trainees are likely to achieve a higher probability of success earlier in training and hence reduce total resources required for effective performance in combat operations. In the long term it may well be possible to specify combat readiness and mission qualification at a quantifiable "proficiency-based" level of analysis in specific skills as opposed to equating readiness to the less accurate and often demonstrably false measure used in current military planning – *experience* measured as total number of hours accumulated in an aircraft and *proficiency* measured by 30/60/90-day "look back" hour and event counts.

REFERENCES

Alluisi, E. (1991). The development of technology for collective training: SIMNET, A case history. *Human Factors*, 33, 343-362.

Bennett, W. Jr., Schreiber, B., & Andrews, D.H. (in press). *Developing competency-based methods for near-real-time assessment in a distributed training environment*. Computers in Human Behavior.

Crane, P. & Bennett, W., Jr., (2002). *The deliberate application of principles of learning and training strategies within DMT*. Paper presented at the NATO SAS-038 working group meeting, Brussels, Belgium.

Colegrove, C. M., & Alliger, G. M. (2002, April). *Mission essential competencies: Defining combat readiness in a novel way*. Paper presented at the SAS-038 NATO working group meeting, Brussels Belgium.

Crane, P., Robbins, R., Bennett, Jr., W. & Bell, H. H. (2001). Mission complexity scoring for distributed mission training. In, *Proceedings of 2001 Industry/Interservice Training Systems Conference*, Orlando, FL: National Security Industrial Association

Edwards, E. (1975). Stress and the airline pilot. In, *The BALPA medicals symposium*. As excerpted from Hawkins, F.H. *Human error, the sources of error, SHEL interfaces*. Human Factors in Flight, Gower Technical Press, 1987.